International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 12, Issue 4 [April. 2023] PP: 129-133

Cardio Monitor: Heart Disease Prediction using KNN

Ritika Dimri[1], Saijal Dahiya[2], Ratnesh Kumar Mishra[3], Vinita[4]
[1]*ABES Engineering College, ritika.19b101011@abes.ac.in*
[2]*ABES Engineering College, saijal.19b101009@abes.ac.in*
[3]*ABES Engineering College, ratnesh.19b101047@abes.ac.in*
[4]*ABES Engineering College, vinita@abes.ac.in*

Abstract. *Every day we here increasing reports of heart attacks and cardiac arrests among people. These heart diseases have become more prevalent and life threatening in the present times. The value for blood pressure, cholesterol or someone's pulse rate is different than everyone else but, we have medically approved values for a healthy heart having person's blood pressure (120/90), Cholesterol (100-129 mg/dL), pulse rate (72), fasting blood Sugar (100 mg/dL), heart rate (60-100 bpm), ECG (normal), width of major vessels (1 inch) in the aorta and (8 μm) in the capillaries. This paper talks of various techniques used for predicting the risk level of each person based on age, gender, blood pressure, cholesterol, pulse rate. Cardio Monitor is system based on predictive modelling and gives results based on data taken as input from the user for various factors affecting them. Cardio Monitor will analyse the values entered by the user and gives the probability of having a heart disease as an output. In this case it is done using KNN after comparing the 5 techniques such as Linear Regression, Decision Tree, Naïve Bayes, Random Forest Algorithm and KNN. They will calculate the probability of the disease and based on their accuracy will let the user know whether they are at risk or not. Average of 92% accuracy was observed.*

Index Terms: *Cardio Monitor, Heart disease prediction, KNN*

Date of Submission: 03-04-2023 Date of acceptance: 15-04-2023

I. Introduction

In our daily lives where heart is such an essential organ as it supplies blood all through the body and makes sure the oxygen supply is maintained through oxygenated blood, in every organ, it also suffers the most. Heart works non-stop and is affected by a lot of factors including our own diet, exercise, stress, and physical activities. Heart problems are increasing nowadays and what makes it worse is that or lifestyle is also making it worse to handle because of the increasing stress, lack of outdoor activities and no work-life balance. The junk food is another not-so-required addition to our troubles to keep out heart safe. Genetic factors are also responsible however if they are identified at the right time, they can be easily controlled. Today, even young adults are losing their lives because of sudden heart attacks. United Nations' World Health Organization reports that 12 million yearly deaths are due to heart diseases alone, worldwide. How these symptoms are felt is also varying for various people, but doctors have fixed set of parameters that they will run a check for whenever any sort of discomfort is felt by patient. Unfortunately, a lot of people are asymptomatic as well and cannot realise until it is too late. The common symptoms include heart palpitations, sudden arm pain and chest pain, difficulty in breathing, cold sweat, and uneasiness. Chest pain is common in many heart diseases which occur due to lack of oxygen supply in organs, called as Angina commonly.

Healthcare is 2nd largest industry of the world economy but importance wise the primary concern of everyone and technological advancements are the need of the hour for them. We need new methods to give information accurately and precisely about chronic diseases so that they can be prevented before worse happens. Advancements on data driven intelligent technologies is disease diagnosis and detection, treatment and research are remarkable. Medical image analysis, symptom-based disease prediction is the part where the most sought-after brains are working. In this paper we aim to present our proposed model on the prediction on diagnosis of cardiovascular diseases with ECG analysis and symptom-based detection. The model aims to be researched and advance in further to become robust and end to end reliable research tool.

II. Literature Review

According to Ordonez, heart disease can be significantly predicted using a few reports from the clients and considering barely 13 attributes such as their gender, pulse, pressure of blood, cholesterol level, etc to say for sure whether a person can suffer from a heart disease or not. In extending their research, they added smoking habits as a factor along with fat/ obesity. Results of data mining techniques such as Neural networks, Naïve Bayes, Decision Trees are applied on the dataset.

Yilmaz, proposed to use least squares SVM, cardiotocogram‘s classification takes place by applying binary decision tree on it.

Noh suggested an associative classifier sort of classification which is built on an efficient FP-growth method. The diverse and huge volume of patterns made them offer a rule to calculate cohesion and allow pruning choices in the pattern-generating processes.

Parthiban have proposed an entirely different and new methodology called as Coactive Neuro-Fuzzy Inference System (CANFIS). This model combines the advantages of neural networks i.e. their adaptivity with the optimisation of genetic algorithm alongside fuzzy to suggest the occurrence of a disease. The accuracies and training set’s performance was noted to suggest that it is indeed useful inn predicting heart related diseases.

Shouman did his work by combining Naïve Bayes with K-means but with different centroid selection for initial step to improvise the Naïve Bayes accuracy. The accuracy in prediction was now 84.5%.

Elma worked with the distance metric and supported the use of K-nearest neighbour and the statistics based Naïve Based classifier as well. The accuracy was 85.92% here.

III. Methodology

3.1 Taking Input

We will take the values from the user for the required attributes. Using Flask API, our data will hit the backend and there our ML model will be presiding.

3.2 Data Collection

First of all, we take 5 techniques to know which one of them is more suited for the task. We considered a UCI Dataset for the same that considered 4 databases. All our techniques were applied on this dataset.

3.3 Data Pre-processing

This is the first step to make the fit for analysing as the data we usually obtain is not dirt-free. A lot of entries might be repeated, or null values may be present that needs dealing with before one can proceed with actual implementation of models. This also increases the tools efficiency and accuracy.

3.4 Predicting with Machine Learning

The aim of this step is to apply all the techniques and find the most accurate one. This section talks about the various techniques we used and the corresponding output they yielded. We chose multiple algorithms suitable for classification purposes.

3.4.1 Logistic Regression

Logistic regression is a classification algorithm where we assign probabilistic values for discrete inputs using a sigmoid function and give the output as probability instead of discrete numbers. The value which we receive can be mapped to multiple discrete classes instead of one (like in linear regression)

3.4.2 Random Forest

It is a supervised learning algorithm. Random forest can be used for both classification and regression problems, here we use it solely for classification problem only. It is randomised but upgraded version of decision tree learning, yielding better accuracy as it takes multiple subsets of dataset to apply the algorithm.

3.4.3 Naïve Bayes

Naive Bayes is a classification algorithm for binary as well as multiple classes. It uses the probabilistic approach as well.

3.4.4 K-Nearest Neighbour

KNN is one of the simplest yet quite accurate form of classification algorithms because it considers some known values in the neighboured with fixed result and then takes the position of the unknown value to calculate the possible outcome. The neighbour most near will affect it most positively.

```
from sklearn.neighbors import KNeighborsClassifier
model = train_model(X_train, Y_train, X_test, Y_test, KNeighborsClassifier)

Train accuracy: 78.10%
Test accuracy: 63.93%
```

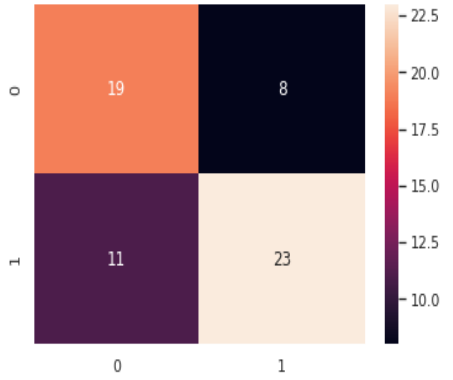

Fig 3.1 KNN confusion matrix

3.4.5 Decision Tree

Here, a binary tree covering all possible outcomes, mostly yes/no in nature, is formed and the leaf nodes cover the outcome whereas label nodes in between cover our possible situations in the environment.

3.5 Integrating UI

Finally, after the model has been selected and applied in the back end, we worked on the frontend using Json files and Flask API to connect the frontend with database to fetch the data entries required.

IV. Results

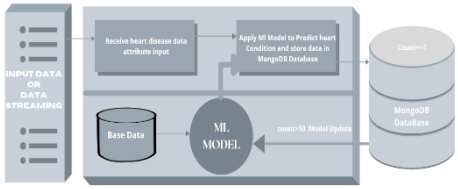

Fig 4.1 Overview of the project

Figure 4.1 shows the overall flow of data when the user gives the input and system starts its predictive calculations.

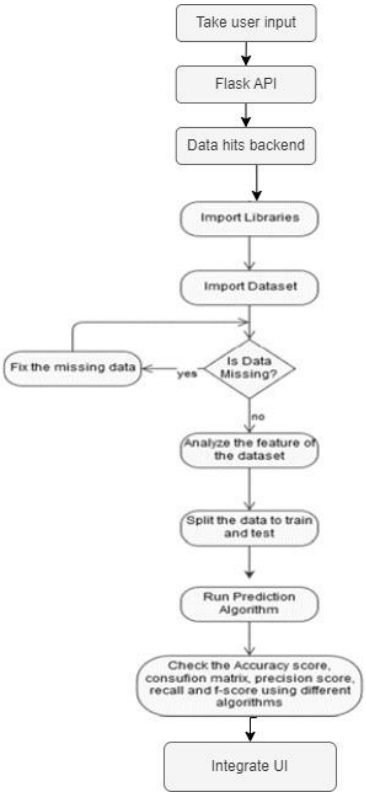

Fig 4.2 Proposed Workflow

Figure 4.2 shows the collected and pre-processed data and various techniques are being applied on the backend with KNN being most suitable for the cause.

Fig 4.3 Final UI after integrating web development

Figure 4.3 shows the interface where the user will enter the input and all the attributes required are clearly asked here with drop-down options, so no confusion is there, and it is easy to use.

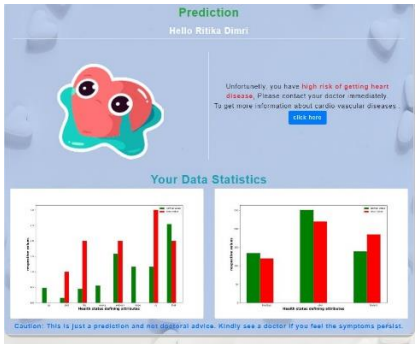

Fig 4.4 Result screen post prediction completion

Figure 4.4 shows the outcome given by the system on taking valid inputs thus helping the patient clearly know the status of their heart and whether they should be seeking doctor's help, or they are fit.

V. Conclusion

The overall objective is to predict as fast and as accurately as possible with least number of attributes taken. Currently, fourteen attributes are taken to predict whether the person is prone to heart issues or not. A lot of attributes are available and reliable, but we aim to reduce the number of attributes required to correctly determine and do it as efficiently as possible. Here, KNN proves to be more accurate and better suited out of the 5 techniques of datamining available. Hence it is most effective way to determine whether a patient should be worried about their heart or not. This project could answer complex queries, each with its own strength with respect to ease of model interpretation, access to detailed information and accuracy.

References

[1]. C. Ordonez, "Improving Heart Disease Prediction using Constrained Association Rules," Tech. Semin. Present. Univ. Tokyo, 2004.
[2]. M. C. and P. M. Franck Le Duff, CristianMunteanb, "Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method," Stud. Health Technol. Inform., vol. Vol. 107, no. 2, p. No. 2, pp. 1256–1259, 2004.
[3]. L. P. and R. Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm," Int. J. Biol. Biomed. Med. Sci., vol. Vol. 3, no. No. 3, pp. 1-8, 2008.
[4]. A. D. and N. R. Niti Guru, "Decision Support System for Heart Disease Diagnosis using Neural Network ," Delhi Bus. Rev., vol. Vol. 8, no. 1, pp. 1–6.
[5]. A. Uma ND, "Extraction of action rules for chronic kidney disease using Naive Bayes classifier.," IEEE Int Conf. Comput Intell. Comput Res, 2016.
[6]. M. S., "Integrating Naive and clustering with a different initial centroid selection methods in the diagnosis of heart disease prediction.," CS IT CSCP, pp. 125–137, 2012.